DOCUMENT RESUME

ED 049 794 LI 002 776

AUTHOR	Work, Robert W.; Phillips, Dennis M.
TITLE	A Study of Textile Information Systems. Final Report.
INSTITUTION	North Carolina State Univ., Raleigh.
SPONS AGENCY	National Bureau of Standards (DOC), Washington, D.C.
PUB DATE	Jun 70
NOTE	40p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Computers, Information Processing, *Information Retrieval, Information Science, *Information Systems, *Search Strategies, *Subject Index Terms, *Time Sharing
IDENTIFIERS	*Textile Information Retrieval Program, TIRP

ABSTRACT

The Textile Information Retrieval Program (TIRP), a study made at the Massachusetts Institute of Technology to develop an interactive information retrieval system operating on a time sharing computer, was demonstrated to and operated by research scientists, information specialists, and numerous other persons at North Carolina State University at Raleigh. The purpose of these trials was to study the interaction of the users with this system and the equipment associated with it and to compare, insofar as was possible, these operations with those of a batch processor retrieval system, using essentially the same data base, operated by the North Carolina Science and Technology Research Center. Approximately 60 searches were conducted in the trials and demonstrations with the cooperation of about 45 representatives of industry and the universities. These trials indicated that TIRP responded more favorably to the simultaneous efforts of information specialists and scientists assisted by expert typists. It is concluded that future developments and extensions should come from the joint efforts of the universities and the textile and fibers industries in order to insure that any subsequent system is both simple in conception and practical in operation. (Author)

ED049794

A STUDY OF TEXTILE INFORMATION SYSTEMS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Final Report

GRANT NBS(G) - 505
U. S. Department of Commerce
National Bureau of Standards

By

Robert W. Work — Principal Investigator

Dennis M. Phillips — Consultant

North Carolina State University
Raleigh, North Carolina

June 1970

LI 002 776

ED049794

ABSTRACT

The Textile Information Retrieval Program (TIRP), a study made at the Massachusetts Institute of Technology to develop an interactive information retrieval system operating on a time sharing computer, was demonstrated to and operated by research scientists, information specialists, and numerous other persons at North Carolina State University at Raleigh. The purpose of these trials was to study the interaction of the users with this system and the equipment associated with it and to compare, insofar as was possible, these operations with those of a batch processor retrieval system, using essentially the same data base, operated by the North Carolina Science and Technology Research Center. Approximately 60 searches were conducted in the trials and demonstrations with the cooperation of about 45 representatives of industry and the Universities.

These trials indicated that TIRP responded more favorably to the simultaneous efforts of information specialists and scientists assisted by expert typists. It is concluded that future developments and extensions should come from the joint efforts of the Universities and the textile and fibers industries in order to insure that any subsequent system is both simple in conception and practical in operation.

KEY TERMS

Information retrieval, information retrieval effectiveness, information systems, search structuring, subject indexing, textiles, fibers.

ACKNOWLEDGEMENTS

The authors wishes to acknowledge contributions of the funds, time, effort, advice and support of every kind that made this study possible. The following organizations and persons are especially thanked:

A. Financial Support

National Bureau of Standards, U. S. Department of Commerce

Allied Chemical Corporation
Burlington Industries, Inc.
Celanese Fibers Company
Cone Mills Corporation
Dan River Mills
Fabric Research Laboratories
Kendall Textiles Division
J. P. Stevens and Company, Inc.

B. Other Support and Cooperative Effort

American Enka Corporation
Miss Alice Laubach
Beaunit Fibers, Inc.
Miss Betty M. Sanborn
Burlington Industries, Inc.
Mrs. Darlene L. Ball
Call-A-Computer
Mr. Robert Patton
Mr. Vern Rodberg
Center for Occupational Education
Dr. John Koster
Celanese Fibers Company
Miss Helen G. Sommar
Cone Mills, Inc.
Miss Audry Anthony
Miss Jane McNairy
Dan River Mills
Miss Hilda Shelton
Hercules Fibers, Inc.
Miss Marjorie Ohl
IBM Sales
Mr. Leon Gentry
Institute of Textile Technology
Miss Nancy Killian
Mr. Dennis Loy
Mr. Robert S. Merkel
Manchester Institute of Science and Technology
Dr. J. W. S. Hearle
Massachusetts Institute of Technology
Prof. Stanley Backer
Dr. Roger Roach
Dr. Robert C. Shelton
National Bureau of Standards, Department of Commerce
Dr. F. Cecil Brenner
North Carolina Science and Technology Research Center
Mr. Peter J. Chenery, III
Mr. Arthur Lockwood
Mrs. Nancy Swyka
Mrs. Mary Ann Williamson
Southern Bell Telephone Company
Mr. John Moore
University of North Carolina at Chapel Hill
Miss Dora Lynn Hickey
North Carolina State University at Raleigh
Mr. James G. Baker
Dean David W. Chaney
Dean Ralph W. Cummings
Prof. Edmund P. Dandridge, Jr.
Prof. James K. Farrell
Prof. Richard D. Gilbert
Prof. George Goldfinger
Mr. Perry L. Grady
Prof. Thomas H. Guion
Mr. Everett H. Hinton
Prof. David H. Howells
Dean E. Walton Jones
Prof. Paul E. Lewis
Dr. I. T. Littleton
Mr. William C. Lowe
Dr. Leroy B. Martin, Jr.
Mr. Rick Porter

North Carolina State University at Raleigh (Continued)
Prof. Henry A. Rutherford
Prof. F. O. Smetana
Mr. Robert Smith
Prof. M. H. Theil
Mr. Manfred Wentz
Prof. Carl F. Zorowski

I. OBJECTIVE -

The objective of this study was to provide a first step toward the establishment of a Textile Information Center. (In the Research Triangle, North Carolina area)

II. INTRODUCTION -

The Textile and Apparel Technology Center (TATC) of the National Bureau of Standards (NBS) of the United States Department of Commerce initiated activity in textile information problems in 1965. At that time a project was put underway to develop a textile information storage and retrieval system under the direction of Professor Stanley Backer at the Massachusetts Institute of Technology (MIT). Research on this subject has continued there up to the present time. This has been the starting point and in large part, the justification, for the study made under this contract, NBS(G)505.

Based upon the activity of M.I.T. three reports (1,2,3) have been issued, four papers have been published (4,5,6,7) and two editions (8,9) of a thesaurus have been issued. A number of oral presentations have been made at formal meetings of scientific groups, as for example, The Fiber Society, Spring Meeting, May 4-5, 1967, in Asheville, North Carolina, and informal gatherings of people concerned with information systems. Most importantly, the system has been demonstrated by Professor Backer and his co-workers. Its operation was first shown at the annual meeting of the Textile Research Institute in April, 1967 and a transatlantic telephone connection made possible a similar showing at Manchester University on July 7, 1967. Another was held under the auspices of the N.B.S. at the Gaithersburg laboratory on September 24, 1969. At this writing another is scheduled to be given at the North Carolina Science and Technology Research Center (STRC) for the benefit of the Textile Industries Information Users Council in May, 1970.

It follows that the entire community that has been concerned with textile information systems and their operation has become well informed concerning the work done and the results obtained at M.I.T. Thus those who have read the publications and especially those who have observed the demonstrations need not be told of the high degree of success of the M.I.T. research and the sophistication that have been achieved in the system developed there.

Although it is generally well known, it is appropriate to restate here that the objective of the investigation at M.I.T. was the development of an operational retrieval system. There never has been any intention to organize a textile information center at M.I.T., with the system so developed being merely a necessary but intermediate step toward such a goal. In the same vein, a data base was produced so as to provide a necessary tool in the structuring of the system. That data base was not intended as a source of information to be used ultimately for retrieval purposes.

But as might be expected, bystanders and participants alike, intrigued and excited by the demonstrated versatility of the system quite quickly raised a not unexpected question. Where does it go from here? It was from this line of thinking that there emerged the concept of a practical examination of the system to determine its potential usefulness in current and day to day information storage and retrieval operations. It was as these ideas began to take form that the seeds of the present trial study were planted in the minds of that community concerned with textile information problems.

Several considerations entered the situation. (1) The primary textile and fiber producing industries are concentrated in the southeastern sector of the country. (2) The largest textile school in the United States is found at North Carolina State University. (3) There are already two information centers operating here. One of these is in the D.H. Hill Library of the University; the other is at S.T.R.C. The former was initiated under the State Technical Services Act; the latter, among other responsibilities, provides information as a regional center for N.A.S.A. (4) Both S.T.R.C. and the University have ready access to computer installations and the highly trained personnel associated therewith.

Discussions among the interested persons led to the initiation of a feasibility study to examine the entire subject. Among the conclusions reached was the observation that a trial study of the M.I.T. system would be a reasonable next step toward the overall long term objectives of the development of a Textile Information Center to be located at some still undetermined point in the Research Triangle Area. It was felt that such a study would provide some guidance as to how and to what degree, if at all, that system could be modified or perhaps used as a reference standard for some local operation. There was no question, at the time, of even attempting to make a transfer of the M.I.T. system as a package for the very simple reason that such would be impractical.

For those who may not be informed of the reason for this last conclusion, it is explained here. The M.I.T. Textile Information Retrieval Program (TIRP) operates within the context of a Compatible Time Sharing System (CTSS) resident in an I.B.M. 7094 computer. There is no equivalent computer in the Research Triangle area, those available being of the I.B.M. 360 series. The latter do not have the CTSS function and TIRP, as written, is therefore not compatible to them. At the time when this particular matter was under examination a rough estimate of the cost of rewriting the program to make it compatible was around $50,000. This estimate, however, was rather casual and would certainly be open to revision.

The aforementioned feasibility study had come forward with the opinion that the entire next step might be expected to cost about $83,000. It was the conclusion that this would be a reasonable investment to be added upon the foundation of the $271,000 expenditure

(11) already committed at M.I.T. An examination of sources of funds including Federal and North Carolina agencies, as well as potential industrial sponsors led to the finding that there simply was not that amount of money available at that time or would be in the foreseeable future for such an enterprise.

To add to the uncertainties, word came that the 7094 computer was to be phased out at M.I.T. in the (then) near future and upon this occurance any possibility of a trial would disappear. It was at such a time and in such an atmosphere that the proposal and program upon which the present contract is based came into being. It appeared that NBS might be able to provide funds in the amount of $10,000 if other sources would match this amount with an additional $5,000. After receiving encouragement from representatives of industry an appropriate proposal was prepared and processed through the usual channels.

III. THE PROGRAM: A BRIEF REVIEW OF PROGRESS AND PROBLEMS

It was visualized in the summer of 1968, when this project was evolved, that provision for practical trials of the M.I.T. system by individuals or small groups from a console located at North Carolina State University (N.C.S.U.) School of Textiles would be made. Such trials would be based upon "real world" problems, but of necessity they would be required to stay within the confines of the M.I.T. data base. Concurrently, as nearly as would be feasible, the same problems would be submitted for processing upon a local computer. As has already been mentioned, since the two systems would be incompatible, this would require that an appropriate program be found, adapted or devised to make such a comparison possible and furthermore that the M.I.T. data base be transferred to the local computer.

The entire situation looked reasonably straightforward to all of those concerned in the Research Triangle area, at M.I.T., the N.B.S. and among the members of the industry which would contribute the funds. The proposal already mentioned was drawn up and processed in August of 1968. It was in the amount of $15,000, $10,000 of which was to be secured from N.B.S. and the remainder from industry and the study was to be completed within a timeframe of 90 days.

The matter of 90 days took on an added significance when it was learned that the normal delivery period of the I.B.M. 1050 console needed for the operation was also 90 days. If the completion of negotiations had been awaited the contract would expire before the equipment would be on hand. Also there was the ever present thought that the I.B.M. 7094 was to be phased out at some unknown but perhaps imminent date. A decision was reached, the I.B.M. people were pressed, they cooperated handsomely and the console was installed on November 11, 1968.

Unfortunately, after what appeared to be an auspicious beginning, a long series of difficulties and delays occurred. At that time the situation was most disappointing to all concerned but with the advantages of hindsight some benefits can be seen to have come of it. The authors of this report and their associates were forced to improvise and in so doing, work out some of the problems which otherwise would have faced them in the trials as originally planned.

It so happened that one of us (RWW) was the chief investigator on a research project, one part of which included the preparation of a bibliography covering a very limited area of fiber preparation. This lacked an adequate index and when the contract was being renegotiated, coordinate indexing and conversion of terms and subject matter to a computer based system were included in the revised version. The authors wish to take this occasion to express their appreciation to the monitor, contract officer and other concerned parties of the Air Force Materials Laboratory, Wright Patterson Air Force Base, Ohio for their cooperation in this connection.

The bibliography was successfully indexed using the aforementioned M.I.T. Thesaurus. Concurrently, using personnel and facilities made freely available through the cooperation of Mr. Peter J. Chenery of S.T.R.C., a program and system was made available which made use of the data base thus produced. Although not a duplicate of the C.T.S.S. at M.I.T. this system allowed an operator at the School console to "converse" with the I.B.M. 360/75, performing real time retrieval operations. Considerable experience was gained.

Additionally, other steps were taken to prepare for the trials and at the same time make use of the equipment rented for that purpose. These need not be detailed here. Murphy's law, "If anything can go wrong with equipment, it will" was reaffirmed. Problems with the console, the telephone and associated switching operations and the computer itself were faced and solved. Most importantly, some of the human problems became quite apparent. Indeed, it was at this stage that the senior investigator (RWW) concluded that he should directly involve himself in such details as typing at the console, thus making himself quite simply "part of the problem". Specifically it was learned, among other things, how exacting a computer can be, how easy it is to make typing and other errors that tend to produce an impasse, how quickly a human being comes to blame a computer for supposed wrongs and how utterly frustrated one can become when the computer is without question in some kind of bind.

On Thursday, September 4, 1969 the material then available for the M.I.T. data base was loaded into the I.B.M. 7094 computer and at 9:55 P.M. conversation started from the I.B.M. 1050 console at North Carolina State University. In the days that followed, the system was used to gain experience with problems suggested by members of the faculty of the School of Textiles. On Friday, September 19 the full data base was available on the M.I.T. computer and the first trial

was made with members of industry immediately thereafter. The work done during this period will be discussed in a later section.

It developed that Professor Backer faced the need to use the M.I.T. computer space for work associated with the translation of the thesaurus into several foreign languages. Since a deadline was imposed by the date of the meeting on the subject to be held in Barcelona in October, it was necessary to unload the M.I.T. computer of the material then being used in the trials. In discussions that followed, it was agreed that situations at both institutions and affecting the several parties concerned, would make it necessary to hold the program in abeyance until after January 1, 1970. In the interest of conserving funds and in view of past uncertainties it seemed best to terminate the rental of the I.B.M. 1050 console and the related telephone installation in the School. Through the kindness of Mr. Vernon G. Rodberg of "Call-a-Computer" Raleigh, an ASR33 teletype and adequate working area was made available for the first trials with representatives of industry on Tuesday, February 3, 1970. This equipment was later transferred to the office of the chief investigator and most of the remaining trial work was done there. The results of these trials will be reported upon in a later section. These trials ended at 4:225 P.M. on Friday, February 13, 1970.

As had been mentioned earlier the original plan called for problems to be submitted to the S.T.R.C. batch system concurrently with their being run on a conversational mode with the M.I.T. system. Experience showed that this was not possible. After the first couple of days it became apparent that whatever planning was done for any given trial, when such a trial got underway, the plans changed, as indeed had been one of the main considerations during the development of the M.I.T. system. Therefore the problems submitted to S.T.R.C. were rather a posteriori and accordingly contributed an additional variable to an already unwieldly situation. The details will be reported in a later section.

IV. THE TRIALS -

A. General

Before the actual trials were started in September of 1969, it had been visualized that it would be possible to make a rather uncomplicated examination of two information retrieval systems. With the first of these, T.I.R.P., a direct interface between the researcher or information specialist and the M.I.T. computer would be possible. It would operate in a conversational mode, the searcher asking questions, the computer supplying answers, which in turn would lead to further questions and answers until the searcher reached his objective. One or another of the investigators under this contract would be present essentially to observe but also to provide varying amounts of preliminary instruction, since this matter of amount of training would be one of the variables to be studied. It was assumed

that some mistakes and incorrect commands would be made by the searcher during the conversation but that these would be corrected, preferably by responses from the computer itself, but if necessary, by the observer. It was not considered that the trials would require the services of a highly trained intermediary who would interpret the instructions of the searcher to the computer and operate the console. On the other hand the use of such a person was visualized in some trials, for purposes of determining relative efficiencies.

In the case of the second system, STRC-IVS (Inverted Search System) a quite different set of trials were planned. These trials were to be conducted in the normal retrieval environment of STRC and direct interplay between the machine and searcher would not be possible. A more conventional batch system would be the basis of trials. An intermediary would be necessary so that the needs of the searcher could be placed into a form for submittal to the computer for processing. Although not normal practice, it was considered that special arrangements could be made at STRC so as to process the input from these trials on a high priority basis.

In the period during which delays prevented the start of the trials on the TIRP system it became apparent that these early plans were much over simplified. Changing circumstances, the development of unexpected situations and the reactions of human beings all combined to force changes in plans and placed a premium on quick adaptability. In any case, a trial of any information retrieval system, whether it be real-time interactive or otherwise, must be approached with considerable care. This is because, simply if one would presume to set up an environment in which he would conduct an evaluation of a technique with which potential users are unfamiliar, he must begin at the beginning and procede in a straight forward manner. This naturally implies a certain demonstration of operating techniques and display of the more elementary procedures before one proceeds to use and develop expertise with the new sophisticated capabilities of the system. Few of the attendees at these trials had previous experience with real-time retrieval systems, although they all were familiar with the basic problems of information technology. The people working on these trials had various levels of experience in the field - some had none, and others had considerable depth and experience. Therefore the task of demonstration and trial involved education of the user to the device, and in some cases, and introduction to the practical aspects of information retrieval.

Overlaying all of these aspects of the trials was the concern for budgetary limitations; and, the fact that since most of the potential cooperators were not located in the immediate area, the time that they could spend with the system was limited. Consequently, the interest was to compress as much experience with the system as was possible into a rather restricted time/money frame.

Prior to each session with the console, attendance at each being from two to about six people, a short discussion was held in which the

basic operational characteristics and command structure were discussed. These discussions included description of the connection between the console and the computer and some elementary aspects of the time-sharing environment of CTSS and how TIRP operated within this environment. It was brought out that these trials were concerned with a study of the interactions between people, equipment and systems, with some hope of learning something concerning the relative costs and efficiencies of various combinations of these in information retrieval operations. It was pointed out that only a small part of the total system might be utilized in any one of these tests. The primary emphasis was upon the search components of TIRP, and while this was a very important part of the system, it by no means represents the complete system. Those persons who were interested in other phases of TIRP were referred to Professor Backer's papers (4,5,6) on the systems.

In a parallel vein, this situation also affected the level of command structure which was used in the trials. Because it was necessary to present a basic background in information retrieval to certain of the attendees, the more straightforward of the command structures were used to begin the demonstrations. These commands, generally, were those which retrieved document strings which were indexed under one keyword or Boolean combination of keywords. The point was made that the result of such commands was to create sub-files which contained entries related only to the keyword or logic equation in use at that time; and that these files were saved so that they could be recalled at a later time. It developed that the primary interest of the users generally fell into this particular mode of operation - their principle concern was with document counts (hits) and with certain bibliographic data in the system relative to those documents. Of secondary interest, it seems, were the more sophisticated commands which provided classification of document strings by category and author. A typical search would progress to the point at which a reasonable number of documents, usually less than 20, could be listed and their titles, authors, journals and associated data would be printed out. On certain occasions the flexibility of the system which permitted one to use the author's name as a keyterm attracted some interest, and classification of a file by author name was used to indicate which author had contributed to a particular collection. Because this program was directed towards a users' trial rather than towards a demonstration of the more sophisticated capabilities of the system the authors of this report did not attempt to push the users towards these commands. Of this aspect more will be said later. But in this connection it may be mentioned that in general only members of the faculty tended to pursue information through the use of authors' names. It is true that upon occasion an information specialist searched tenaciously for a given document by a given author, having known from another source of its existence. But this was, of course, a check upon the completeness of the data base, rather than a trial of the system, and was not truly within the scope of this investigation.

One particular aspect of the system's capability which was emphasized heavily was the fact that the M.I.T. Thesaurus of Textile Terms was the basis and guide for selection of keyterms. All files had been constructed with this thesaurus as a check against invalid terms, and the search module used the thesaurus as a means of access to those files. The command to list the hierarchial and related term structure of a given keyterm found considerable use. This was particularly noticeable when a user had selected a term with which he was familiar and had suspected had been used in the indexing of the file, only to find that, in fact, there were no entries listed under the keyterm in question. At this juncture, the thesaurus was brought forward, either through the console or through examination of the printed copy, and the proper term for indexing was given.

As a passing point, some of the searches received more preparation than did others prior to operation of the console. This was in an attempt to investigate the learning rate and assistance which the system gave to the user in the event of no previous study of applicable terms. An alphabetical listing of authorized keyterms, with the number of postings of each entry had been prepared from a copy of the M.I.T. file which had been loaded at STRC. This listing was not a thesaurus, but a print-out of authorized terms taken directly from the file, and contained none of the structure provided by the thesaurus. Some attendees were encouraged to use this listing as a study aid prior to initiating a search as an alternate means of selecting keyterms and to enable the operator to determine the number of postings to be found under those keyterms in the data base to be used. As an alternative to this, the operator would enter the machine with a simple logic query merely to determine the number of entries available. Of course, this might or might not require a considerable amount of computer time, depending upon the number of entries; but in those cases where it was expected that a very large (>1000) number of postings was to be found, the STRC dictionary was consulted first. This was done as a means of conserving both computer and line time. It developed that in many cases the users preferred to consult this list of postings rather than use the thesaurus either in its printed form or from the console. This could be attributed to the concern of many of the users with the number of entries which might be retrieved. This emphasis upon the number of hits in response to a query, as opposed, say, to emphasis upon selection of the most relevant document among those presented, poses some interesting points. The TIRP search module contains many sophisticated techniques, the intent of which is to enable the machine to find, or to assist the operator to find, the most relevant document(s) within the file or the author of these documents. The approach most usually taken was to use the machine in a direct manner to obtain a document list of reasonable length, and then to perform value judgement concerning relevance by examination of titles, authors, and keyterms. This aspect will be further commented upon in a later section.

B. Discussion

As has been mentioned earlier these trials were aimed at providing for the study of the interactions between people, machines and systems. The investigators found it all too easy to allow themselves to move in the direction of demonstrating, and, indeed, when the audience was composed of University or other administrators, this was the logical approach and was used. But it was learned that it was not an easy task to involve research people in the trials. Many who at first thought would appear to be the ones who would take an interest, simply were not concerned. One such who bothered to give an explanation, said in essence, "There is only one proper place to make a literature search and that is in a good library. And there is only one good way to do it and that is, go to the library and start reading."

Yet others, attracted by the idea of new and better approaches to literature searching were prepared to watch and even admire the operation of the communication linkage, but could not be brought to go beyond that point. To some it seemed to offer future promise but, they explained were quite busy currently and really had no immediate literature problems. One person was prepared to submit a list of keyterms with the request that he receive a printout of material on every document in the data base locatable by the use of these key - terms. He then planned to assign a graduate student to cull through these and select the ones of interest.

The great majority of literature users agreed quite readily that computer based systems of searching would at some date in the near future become generally necessary and commonly used. But more often than not there was a great reluctance to concern themselves with the here and now. Almost without exception there was a desire to make use of an intermediary who would be expected to take the assignment and return with the answers. Little concern and less interest was shown in how these answers would be secured. It hardly need be said that the investigators came away somewhat shaken from some of these encounters. But it was not the objective of this study to perform searches merely to secure a limited bibliography for a requester. It was insisted that the client be present during the study.

All of this is mentioned not to bemoan the lack of concern of the scientific public, or at least one small cross section of it, in a study close to the hearts of the investigators. Rather, it is to emphasize to the reader of this report that those persons who become active participants in the trials were the "exceptions to the rule" rather than being the average or random researcher. Having learned this early in the study, the trials of February 1970 depended heavily upon the interest of professional librarians and information specialists.

With this background in mind, let us examine the trials started in September, 1969.

Several members of the faculty of the School had evidenced a desire to work with the investigators. Of these, two were working in research fields not covered by the data base. One was concerned with dielectric and acoustic behavior of fibers and the other with chain folding and spherulite formation in linear, fiber forming polymers. Searches in these subject areas were doomed to be unproductive since they were not included in the data base. Nevertheless their curiosity was not to be easily dismissed and searches were made. It goes without saying that the investigators gained more in experience in these cases than did the researchers in securing specialized information.

The needs of the others were more closely associated with the data base, being concerned with dyeing properties of fibers and the physical behavior of end products. In these subject areas some material was retrieved of value to the participants. In general, searches were made for all comers, no matter how forlorn were the hopes of securing vital information from the data base available.

C. The Operations

As has already been mentioned some members of the faculty did not concern themselves with the trials or indicated they did not see that they had any need for information retrieval searches from a limited data base. It developed that such views occurred most frequently among the senior members of the faculty. While the first reaction of the present authors was to classify the response as being based on academic conservatism, a little more thought led to the conclusion that such was not the case. In truth these men were already well acquainted with the published literature in their selected fields and kept constantly up to date with it. It was the more junior members and especially the graduate students who were more eager to cooperate in making trials.

This situation played an important part in the use of the M.I.T. system and data base. The scientist who knew the published literature, the titles, the authors and the sources and could make use of all of these in pursuing and tracking down more or less obscure information had no need to do so. On the other hand, those of far less experience had insufficient background to engage in such tactics. It even happened that when one of the former group was induced to make a trial,he might become impatient as papers by himself and others who he knew were listed. Granted that the system easily could have been used to exclude the searchers' own publications and those of authors well known to him. But it is believed that few human beings will act in such a manner. Conversely, the younger man wanted a complete list of papers and he wanted the titles, authors and citations and, almost without exception, nothing else.

The usual individual researcher was found to be but little concerned with the structure and potentialities of the information storage and retrieval system, per se. Instead of being prepared to listen to the long and, indeed, involved presentation of the description of

what could be done, he wanted to see some results. Generally speaking this resulted in an attack via the use of keyterms, since the concept of the associational train made possible thereby is straightforward, quick to be grasped, and can be responded to almost instantly. Thus it followed, more often than not, that a single essential and rather broad keyterm would be selected and used to formulate the first command; 1/

X = Keyterm A

When the response was printed out, a decision was called for. If the number of documents given was about 20 or more, as was usually the case, since the first keyterm was broad, a second keyterm would be selected and put into the command;

Y = Keyterm B

On the other hand, if there were but a few documents in X, then the second request would be aimed at securing detailed information on these. But assuming that two keyterms, based on X and Y, were used, the most common next step was to perform an intersection between these by means of the command;

Z = .X * .Y

This usually put the searcher into a position where the number of documents faced were of a sufficiently small number that he would be prepared to cope with the details concerning them. He then would request a printout to secure most or all of these details by means of an appropriate command sequence.

In addition to the trials made with members of the faculty and graduate students cooperating during the period of operation in September, 1969, several short demonstrations were made for the benefit of administrators and other persons who expressed an interest. Since the objective of these was to exhibit the operations, rather than perform studies, they need not be reported upon here.

As has been mentioned earlier in this report, at the start of the operations made in September the expected and full data base was yet to be loaded. For this reason, trials involving representatives of industry had been deferred until it would become available. (Perhaps by way of illustration it may be mentioned parenthetically that the use of the keyterm "cotton" before the full data base was loaded drew a response of "402 documents", after, exactly twice that

---

1/ A tabulation of commands used appears in Appendix I.

many i.e., "804 documents") When the complete data base became available, circumstances prevented its immediate use, as already recounted and thus trials with members of industry took place but two days. It seems best to discuss these in conjunction with the trials made in February, 1970, since both generally involved representatives of industry who were information specialists rather than being persons directly engaged in research. But before describing them it seems appropriate to mention the work done in that period with another group, neither bench researchers nor information specialists. This was to inform administrators, almost entirely limited to the University and associated technological community, since this investigation had as its objective making a first step toward the establishment of a Textile Information Center in the Research Triangle area. It has been already stated that trials which would involve others in direct operations could not be achieved with administrators, for reasons that are obvious. Thus, demonstrations, rather than trials, became the order of things. These tended to emphasize the versatility of the TIRP and had little to do with the solving of real world textile information problems, except that wherever possible subject areas were used which might be of interest to the observers. It is the opinion of the authors of this report that satisfactory performances were given and that the potentialities of TIRP were impressed upon the onlookers. The commands used in these demonstrations have not been included in the tabulations and graphs given in the appendices.

Most of the trials made in February, 1970 were made with the cooperation of representatives of the textile and man-made fiber industries. Certain behavioral patterns on the part of those persons directly engaged in research have already been described. It was quickly observed that these differed materially, one may say in kind rather than in degree, from those of the information specialists. This last group were found to be deeply concerned with the operation of the system and quite well acquainted with the papers already published describing it. They came to the trials with specific information retrieval problems in hand, in many cases these problems had been worked over with their own facilities. Without exception they were open minded and receptive.

It is not unexpected that information specialists should have a quite different approach to these trials than did researchers. In the organizations that the former represent, the latter are the clients. The former are the intermediaries between the computers and the clients, since in no case is there a direct connection and a time-shared operation available for the use of the researcher.

It was said at the end of section IVA that "The approach most usually taken was to use the machine in a direct manner to obtain a document list of reasonable length, and then to perform value judgement concerning relevance by examination of titles, authors, and keyterms". There are several possible reasons which might explain this situation, but it is the opinion of one of us (D.M.P.) that they can be narrowed down to two. (1) The time available to each

user was not sufficient to allow him to become proficient in terms of system command capabilities, and as a consequence, the more sophisticated capabilities were not used. (2) The pattern described above fits directly into that which the majority of these attendees follow in their normal work habits. They have reason to be suspicious of indexers as a result of their experience with other systems and prefer to trust their own judgement in combination with that of the author as exemplified by the choice of the title, and as opposed to that of the indexers and complicated internal machine logic structure, insofar as document relevance is concerned. One got the distinct impression that, once the title and other bibliographic data had been listed, they considered that the machine's function was over, and the only way the operator could feel secure concerning the relevance of documents was to obtain abstracts and document copies. Certain commands automatically throw those documents whose relevance probability is relatively high to the top of the document list and this technique was used to some extent - but the reliance of the user upon this computed relevance was low.

The final results of searches obtained after review and study of vocabulary and logic differed little from those which were obtained with little or no review with one significant exception - the time required to process the searches. Most certainly it was shown that to the degree that the operator had prepared himself prior to addressing the machine, the time required to obtain results which were meaningful to him was shorter. It should be noted, however, that this procedure restricted the operator's flexibility, and caused him to look for significant results at precise stages throughout the search. This is in contrast to the more relaxed operating mode in which the operator came to the machine with only a general idea of the nature of his problem, and with little or no preconceived notions concerning anticipated answers. The people in the latter category were more inclined to use the thesaurus capability and to create more extensive files than those in the former; but in fine, one could not categorically state that the answer produced in one case was superior to that found in another. This has a bearing upon the cost-effectiveness of the system, of course. It is unfortunate that it was impossible to follow the learning curve of an operator all the way through to analyze the techniques that would have been used had any of the attendees been able to develop an in-depth capability on the console. Neither was it possible to give the console to one person with a simple list of unit sections and let him "have a go at it". Time and money limitations would not permit this.

All in all the trials were constructive but it would be a retreat from reality to claim that they were outstandingly successful. In the brief review of the progress of this work some of the delays and difficulties were touched upon. To imply that these had no adverse reaction upon the investigators would incorrectly lead the reader to believe that the authors would suggest themselves possessed of patience and fortitude far beyond the amount that even their most friendly

critics would concede to them. When things went wrong, if only the making of typographical errors or simple mistakes in commands by the person at the console, the operator reacted adversely. The automatic printouts of computer and elapsed times periodically reminded everybody that the former represented a cost of several dollars a minute and the latter a cost of thirty to forty cents a minute. It is perhaps pertinant to mention that both universities and textile companies are at least as sensitive to dollars running away as the next one. And last, but not least, when the computer or some mechanical operation between the observer and the computer went out of order, tempers, and consequently performance of persons, suffered correspondingly.

V. SYSTEMS COMPARISON

A. General

Let us consider two accepted means of information retrieval, with the end results from both, to the person seeking an answer to a problem, being determined by the receipt of a listing of relevant abstracts. It is assumed a priori that the abstracts do, in fact, describe the true content of the document, and the reader can make a value judgement at this level as to whether or not he needs to read the source document. The first of these is a system such as the time shared M.I.T. - T.I.R.P. and the other is a rapid turn-around batch processed system run in the normal job stream of a computer, not necessarily on site, but within reasonable distance so that communication between search requestor and user does not constitute an overbearing consideration. Assume that the same file content exists in both cases. In the former case, the search requestor or an intermediary obtains the use of a console, and direct themselves to the solution of an information problem in a time-shared environment. Experience has shown that reasonable output, in the form of titles, authors, and the like can be obtained in the space of a few minutes to an hour or thereabout. If the system had contained print-out capability for abstracts, then in the context of this comparison, he would have received the answer to his problem within that time. The costs, which would have included time of personnel, computer, both computational and "hook-up", and telephone line rental would have been directly proportional to the amount of time required, both lapsed and computer, to retrieve answers. The complexity of the problem statement will affect the computer time used.

In the case of the batch system, the search requestor would be required to study his problem using published thesauri, lists of postings, and other pertinent information in order to formulate his problem prior to submission to the computer. He would then fill out certain forms to support personnel for keypunching, and entry into the machine. He must then content himself to wait, using his time to some other advantage, for the answer to be returned to him. This waiting period is familiar to anyone who has ever used a computer, and in most cases is relative and can hardly seem to be as exasperating

as five minutes spent in waiting for a time-shared system to respond. The time lapse in this case is dependent upon variables too numerous to be listed here, but usually it is dependent upon the level of machine usage and the job priority as it affects the system's scheduling algorithm. This level of machine usage also affects the operation of real-time systems as well, in much the same manner. Once the answer has been returned to him, the requestor must make his evaluation, and most probably, decide if another run is required. If such is the case, then he must perform essentially the same steps listed above in order to improve the quality of his answer to an acceptable level. This then, constitutes one of the principal advantages of the time-shared system; for the user of the real time system can modify his problem in a feed-back mode, and direct its progress in light of answers obtained at intermediate steps. The opportunity to modify a problem, and to direct the solution is also available to the user of a batched system, but the total time required between problem initiation and completion is considerably greater, as noted above. However, the desired approach is to state the problems sufficiently well the first time so that subsequent runs will not be necessary.

Cost-effectiveness, is in reality, that factor with which information retrieval systems are essentially concerned. One must note that, even in their most sophisticated forms of operation, input and storage, a retrieval system is only a pointer - in only a very few cases can the use of these systems be considered as capable of reaching the ultimate goal itself. The practical objective of any such system is to obtain data concerning papers and documents which are the real sources of information which the operator is trying to reach. Therefore, given the most sophisticated retrieval system which one could imagine, operated by someone intimately familiar with all its capabilities and techniques, it is entirely possible that the results will be useless in the practical sense if (1) indexing fallacies and observations on the part of the author cause the machine to produce documents which are not relevant; or, if (2) once he has received notification that such-and-such paper will, in all probability, satisfy his requirements, the person who initiates the search chooses not to read that paper. So far as this author (D.M.P.) knows, no quantitative data exist concerning this second point, but, from his experience, if one receives a collection of documents, the chances are good that only a fraction-perhaps a small fraction at that-will be studied in depth. Consequently, it behooves the designers of retrieval systems to be pragmatic to a certain extent, and consider the anticipated benefits of the results which will be produced by that system in the strong light of the costs of design, implementation, operation, and maintenance of that system. Insofar as M.I.T.-T.I.R.P. is concerned, one must consider that the project was initiated and carried out as a research effort, and that the end product of that research should not necessarily take its form as a functional retrieval system. It is the result of research about information retrieval systems, and one must not construe this to mean that the project was intended to be used as the basis for the design of a complete information system which one could transfer in total to either an academic or industrial

environment. However, as a sophisticated interactive retrieval system, it is an appropriate model which all can use in an attempt to draw conclusions concerning the relative merits and costs of this class of systems as opposed to others which are in use.

Other than the statistical data concerning the operation of another retrieval system, S.T.R.C.-IVS, which may be found in a later section of this report, one must consider the qualitative aspects as well if one is to draw conclusions concerning cost effectiveness. And, in this light, one would probably come to the position that a strict quantitative comparison is neither possible nor practical; and, furthermore, that the conclusions could be misleading. Of these considerations, such things as urgency, critical need and others could be listed but one must consider in some detail the so-called time-value of the user and the expected returns from the use of his time,in relation to the rest of the system. It is vital to determine the ratio of man-cost/machine-cost at which it becomes practical to provide a particular device or assistance to a particular employee. A parallel can be said to exist when one considers the use of time-shared systems for data reduction and numerical computation required to solve engineering or scientific problems. However, certain very strict limitations exist insofar as this comparison is concerned. It can be said, in general, that most problems of this type which are solved in a time-sharing environment are the one-of-a-kind type - the use of a time-sharing system greatly reduces the time lapse between problem conception and return of data because of the programming capabilities provided by the programming languages and incremental compilers. However, when it is seen that such problems will be run on a production basis, the usual procedure is to transfer the program to a batch processor. Still another consideration is that the results of such computation may constitute an end in themselves - that they are of intrinsic value as they stand. It was pointed out above that this is not usually the case with information retrieval systems. Therefore, one could say that the decision to provide real-time computer assistance could more nearly be based upon quantitative data in the case of computational operations than in the case of retrieval.

B. Specifics

The North Carolina Science and Technology Research Center (S.T.R.C.) is an agency of the state government, and is the operating arm of the Board of Science and Technology. S.T.R.C. is a contractor to the National Aeronautics and Space Administration, and is one of the six Regional Dissemination Centers which provide access for the industrial and academic communities to the technical resources made available as a result of this country's and world's investment in aerospace research and applied science. To accomplish the goal of transferral of aerospace and related technological developments, S.T.R.C. processes the NASA information file in response to requests for data submitted by industry and academic representation. In addition, S.T.R.C. has access to other computer based files, notably

the collections of the Department of Defense, The Institute of Textile Technology, Chemical Abstracts Condensates, and others.

The operation of the S.T.R.C. retrieval system relies heavily upon an interface between the person who requests a literature search and the computer system through Application Engineers acting as intermediaries. Their principle functions are the interpretation of the search question into machine acceptable formatand the screening of search output for relevancy to the search request. It is seen that the search requester then is not directly related to, nor is he principally concerned with the details of computer-based information retrieval - the requestor's primary function is to explain his problem properly to the Applications Engineer. The latter thus has the responsibility, and the means at hand, to comply with the search request. Such an approach requires considerable skill on the part of the Applications Engineer and a reasonably sophisticated clerical and support staff.

In the earlier days of its operation, S.T.R.C. as a NASA Regional Dissemination Center was supplied with the NASA file as well as retrieval programs for using it. The availability of this package made it possible for S.T.R.C. to search the NASA data base without incurring development costs associated with the design and programming of a retrieval system. This advantage was off set to some extent by the fact that the retrieval system was highly machine dependent, and, because of the rigidity of its design, the possibility of processing other files was precluded.

The design and development of the current S.T.R.C.-IVS made it possible to reduce by a considerable amount the direct machine costs required to process searches, and provided a means by which S.T.R.C. could expand its holdings of machine searchable data bases. These additions include the file of the Institute of Textile Technology and the unclassified portion of the Department of Defense file and, most recently, the M.I.T. Textile File.

S.T.R.C.-IVS is an inverted search system which operates in the normal job stream of an I.B.M. OS/360-75. The main retrieval module is written in Fortran IV, and has assembler language sub-routines to facilitate data flow. The principle file design is fixed-blocked ISAM, and the files are currently mounted on 2314 direct access storage devices. Search input is in terms of a Boolean logic statement, with checking functions built in to insure that all subject data conform to file and system requirements.

Search files are of inverted format, but limited bibliographic files are provided in linear format; also mounted on 2314 disks. Retrieval operations are performed using the inverted search files to the Engineers' satisfaction, after which bibliographic citation data are listed. Bibliographic data include full abstract text in the case of some files, and title, author, etc., in the case of the M.I.T. file. Hard copy abstract card files are used in lieu of computer based files

when they are available.

Comparison of the output of the M.I.T. real time searches and those secured by use of the S.T.R.C. system show that essentially the same final hit list was obtained in each case. This is, obviously, a result of the fact that basically the same file was used in both cases. Logic preparation for the batch searches was aided to some extent by using the results of the real time searches as a guide - in some cases as a direct list.

Other searches using both systems separately have been made with essentially the same results. The indication is that thorough preparation using published copies of the thesaurus and lists of postings per keyterm can result in search out-puts which are comparable with those obtained from the console. This situation is, of course, limited to those areas of comparable logic capability and does not extend to the more sophisticated search techniques of the M.I.T. system. It was impossible to extend this comparison analysis to those areas because commands which request searches based upon cited authors, for example, do not exist in the S.T.R.C. system.

Costs of operating the S.T.R.C. system include, naturally, the engineer's time, clerical support time, and computer costs. Computer costs include both the direct search processing time and overhead. As with the M.I.T. system, direct computer processing costs are functions of the amount of data to be processed. These occur in two ways. First of all, computer time is required to process a search request to 'solve' the logic statement in terms of number of documents which satisfy the statement. These computation times are in proportion to the number of postings, i.e., number of documents, which have been indexed under a particular subject term. The second of these is the time required to print the list of document numbers and bibliographic data corresponding to a particular solution. To this extent, most computer-based retrieval system costs are similar; the similarity is reasonably close for the T.I.R.P. and S.T.R.C.-IVS. A considerable degree of dissimilarity exists, however, in terms of the operating systems under which the two programs are executed. In the case of T.I.R.P., which is under control of C.T.S.S, execution is truly on a 'shared' basis with other users. S.T.R.C.-IVS is processed in batch mode, and, although the processing unit may be executing more than one job at any given time, the execution of those jobs is not interrupted until execution is complete. In the case of C.T.S.S. as with other time-sharing systems, execution proceeds on an interrupted basis. The machine performs scheduling functions which call certain jobs for processing, subsequent removal to intermediate storage areas, and recall for continued processing. This results in certain amounts of overhead machine scheduling costs in addition to the normal file maintenance and storage costs mentioned earlier. However, they are a normal part of the time-sharing environment, and must be considered in the design of any system which is intended for execution in that framework.

However, in both systems, the same points made earlier still obtain: search processing time is increased as the number of postings is increased and output time is increased on the volume of output increases. Timing studies made of the S.T.R.C.-IVS show that for practical purposes, the direct time required to process a search can be approximated by the following equation. (12)

$$t = 0.659\ N_T + 0.00456\ N_p$$

where t = time to process
N_T= number of terms
N_p= number of postings

This equation gives the average time, in searches, required to compute search 'answers' as a function of numbers of postings and keyterms. Further studies show that the time required to list the bibliographic dates for M.I.T. file output is of the order of 0.3 seconds per record.

Timing studies taken from these trials are given in the appendices. These data by no means tell the complete story - for in addition to the cost of computer processing time, one must add other charges such as for loading expense, cost of 'hook-up time', telephone rental and long distance line charges and rental charge for console use or other equipment.

There is, however, no need to belabor the point: any attempt to make a comparison of costs between machines of the 7094's generation with those of the S/360-75 would be highly suspect. There are, nevertheless, other ways of approaching the subject. Analysis of T.I.R.P. costs, and of operator reaction, indicates that a majority of time, both computation and elapsed, is consumed by listing of bibliographic data. Consequently, an obvious reduction in cost might be affected by the simple expedient of using a remote (from the operator console) high speed printer to accomplish this task. Furthermore, when one considers the enormous capability of time shared equipment vis-a-vis computation, one is lead to conclude that the most effective utilization of this equipment would be in the area of solution of logic statements strictly in terms of numbers of output documents, and brief bibliographic descriptors such as document category and the like. The intent is to encourage the use of the machine's most effective tool; that of rapid response to an essentially quantitative problem statement.

This leads one to consideration of any number of alternatives. Let us assume that no large volume textual displays would be permitted through the time-sharing mode. This would imply that textual data would then be unavailable to the processing unit, and, as a consequence large data files upon which such data are stored would not be required. Rather, strictly numeric or a limited mix of textual/numeric data would be present. Given that this would require some rather unique approaches

to indexing, document codification, and classifications; such efforts would in effect reverse the trend now extant which causes so much of the machine capacity to be locked down by the problem of data transmittal from processing unit to file storage and back again.

One approach to this problem was published in the appendices to the A. D. Little report, "Documentation and Centralization"(13). A stochastic model of file content was used as a means of predicting output volume in response to a logic statement with some degree of success. This approach is of particular interest in the light of the concern of some of the trial participants over the number of documents (hits) expected as the result of a logic statement.

Such a procedure, using concise data which describes file content in a numeric fashion, as opposed to the standard practice which calls for manipulation of file content itself, could result in a considerable exploitation of the time-sharing computer. In this system, the search files would contain information about the data base of interest. These small files, structured for efficient processing, would be used to determine the most effective strategy and logic statement. Once these strategy statements had been developed, they would be passed to a batch processor which would provide bibliographic data. This approach would be, it seems, the logical response to the two points of interest disclosed by observation of the participants:

1) How much and what is the elementary content of data which one might expect from the solution of a given logic statement?

2) What are the more highly relevant citations which appear as a consequence of solution of that logic statement?

VI. COOPERATORS' COMMENTS -

Consideration has been given to the relative desirability of quoting letters in their entirety or attempting to summarize and consolidate them with the oral comments which were received. The latter plan has been selected. The thoughts that follow are not in order of priority or emphasis and if injustice has been done to the opinions of any commentator, the authers must accept the onus of blame for so doing.

1. The trials could have been better constructed and it would have been desirable to have had them completed at an earlier date.

2. The common opinion was expressed that the combination of qualified typist, expert information specialist and inquiring bench scientist commonly does not and probably should not be expected to exist simultaneously. Therefore, if a retrieval system, operating in a conversational mode is to be used successfully, some compromises must correspondingly be developed to render it effective.

3. There was criticism of the programming and the method and content of the information returned to the searcher. Some representative comments, for purposes of illustration follow:

A. On an "UPRINT X" command the document number, an English translation of a title and the author are returned to the searcher. Additionally it was desired to know, (1) whether the language of the original document is English and (2) the abstract source and description in accepted abreviated language, rather than in code.

B. There is no simple and single command which secures only the most wanted information related to a document, in the order familiarly used by searcher and scientist alike. This is, (1) title, (2) author, (3) journal or patent (4) year or volume number and (5) pages. It was felt that this set should be directly obtainable from either a document number or the contents of a file previously accumulated.

C. More attention should be paid to volunteering the language of the original document and the availability of an English summary.

D. When an illigitimate keyterm is used and the computer responds with "near misses", none of these should be additionally unacceptable and thus require a second inquiry for the searcher to receive the instruction, "use _____". One operation should suffice, especially since these responses consume a considerable amount of both computer and elapsed time.

E. It was pointed out that every adult who has reached a place where he or she has need to organize a literature search, has already become expert in the use of dictionaries and phone books. It is, therefore, no problem whatsoever to learn to use a thesaurus or

alphabetical printout of postings available in the data base and the number of each therein. In turn, the location of a keyterm, its legitimacy, its place in the hierarchial structure, a run through of associated keyterms and the number of postings to it currently in the data base occupies the searcher only a matter of seconds. The effort, time and cost of doing the same via typed question, computer operation and printout would seem to eliminate the need for the use of this route.

F. Unless care is used in some of the command system to use "no dup", duplicate information is normally printed. It was felt that the operation should be "fail safe" in that duplication should occur only when requested. Indeed, in trials described, even with the use of the "no dup" command some titles were printed out repetitiously in association with each of the joint authors, and then each of these repeated as many as four times.

G. The need for printout of the reference number of keyterms, unless requested or needed for some subsequent usage, is a waste of time and effort.

4. The need for meticulous proof reading and debugging of input was emphasized. For example, in one case a cooperator had seven hits in a response. The names of the authors of three of these were misspelled and the language of one was incorrectly stated.

5. There was general agreement that the entire TIRP operation has made a substantial contribution to knowledge in the field of information storage and retrieval, especially from the standpoint of future potentials. These trials were seen as a reasonable part of that program.

6. It was considered to be unfortunate that a situation had developed which prevented the use of the full and versatile command structure throughout the trials.

VII. CONCLUSIONS (AUTHORS')

1. TIRP, as developed by Professor Stanley Backer and his associates at M.I.T.: the fundamental logic upon which it has been based, the thesaurus and its hierarchy structure, the information storage and retrieval system, constitute a record of outstanding achievement.

2. The versatility of the system allows for virtuosity on the part of the initiated user and almost limitless approaches to computer based information retrieval.

3. The trials described in this report, limited though they were, have led to some practical results which may be found to be useful by those persons who are concerned with future applications of TIRP.

A. The person conversing with the computer via console or other typing device, to be fully effective, should possess the manual dexterity and motor skills to permit him to type with speed and accuracy.

B. This person should have knowledge of the logic capabilities and be familiar with the bases of the system and the related tools and their operations. He should know how the computer system is constituted and the computer response to commands.

C. He should be widely versed in the literature of the specialized field in which he is making search. Ideally, he should be the research scientist who is desirous of securing information for his own needs.

D. It will be extremely difficult for one person to fulfill the conditions stated in items A, B and C above for the following reasons:

1) The skills of item A are those of a machine operator.

2) The skills of item B are those of an information specialist.

3) The skills of item C are those of a scientist.

E. To the degree that conditions A, B and C are not met simultaneously, the overall operation may be inefficient or needlessly expensive or both.

4. It would have been distinctly advantageous if it had been possible to make these trials during the development of TIRP, for feedback purposes, rather than after its completion.

VIII. RECOMMENDATIONS (AUTHORS')

A. Some organization, and the Textile Industries Information Users Council appears to be the most appropriate at this writing, should pick up responsibility for the future study of the practical usefulness of TIRP. Among other aspects, there is the need to re-write the programs used on the IBM 7094 for use on other computers, most especially, of course, those which will support a conversational mode operation.

B. Industrial information operations offer the best opportunity for developing the potentialities of TIRP. They provide a framework in which it is possible to maintain controls and the associated disciplines necessary to determine aspects such as cost effectiveness.

C. If and when actions may be taken along the lines suggested in paragraphs A and B above, it will probably be advantageous to restudy the command system with an aim toward simplification.

D. Consider the development of a several step retrieval operation of which the essential elements are:

1) Expand and modify the MIT thesaurus as experience and usage dictate, maintaining the existing hierarchical structure. (Some organization such as the Users Council might accept this responsibility).

2) As an adjunct to the thesaurus, provide a regularly updated printout, in dictionary form, of the number of postings against each acceptable keyterm. This should be used by searchers in conjunction with the thesaurus.

3) Adapt those functions of TIRP where conversational mode response is available, to enable the user to narrow the confines of his search problem rapidly.

4) Provide for a delayed printing of specific information such as title, author, journal (or other), date, language of original document, source and description of abstract and notation as to whether the journal is located in the house library used by the searcher.

5) If item 4 is handled by a central information center, remotely located to the searcher, some system should be worked out to tell him where he may conveniently locate journals, patents, abstracts and other information to which the computer has referred.

6) In the development of a practical information storage and retrieval system, there should be a parallel study of its operation for feedback purposes.

BIBLIOGRAPHY

1. Design of Information Storage and Retrieval System: Part 1. Patterns of Flow of Technical Information: A Study and System Design Problem for the Textile Industry. Richard D. Wells and Stanley Backer. January 1967. U. S. Clearinghouse Document P.B. 176,549.

2. Design of Information Storage and Retrieval System: Part 2. Design of an Information Storage and Retrieval System for Textile Literature. Attila T. Alptekin and Aiden B. McNamara. January 1967. U. S. Clearinghouse Document P.B. 176,550.

3. Design of Information Storage and Retrieval System: Part 3. Development of a Time-Shared Storage and Retrival System. Robert C. Sheldon and Stanley Backer. January 1967. U. S. Clearinghouse Document P.B. 176,551.

4. The Problems of Textile Information Retrieval. Stanley Backer, E. I. Valko and M. Liang. Textile Research Journal. 37: 880-894 (1967).

5. The Thesaurus as a First Step in an Information-Retrieval System. Stanley Backer. The Textile Institute and Industry. 5: 91-96 (1967).

6. Design of an On-Line Computer-Based Textile Information Retrieval System. R. C. Sheldon, R. A. Roach and Stanley Backer. Textile Research Journal. 38: 81-100 (1968).

7. The Mather Lecture 1969: Textile Information Storage and Retrieval in the 1970's - the Role of the Professional Societies. A. B. McNamara. The Textile Institute and Industry. 7: 209-212 (1969).

8. Thesaurus of Textile Engineering Terms. First Edition. Stanley Backer et al. Massachusetts Institute of Technology, Cambridge, Massachusetts. February 1966.

9. Thesaurus of Textile Terms. Second Edition. Stanley Backer and Emery I. Valko, Editors. The M. I. T. Press, Cambridge, Massachusetts and London, England. February 1969.

10. The M.I.T. Textile Information Retrieval Program. William C. Lowe, Robert Chartrand and Dennis M. Phillips. School of Textiles, N. C. State University, Raleigh, N. C. March 1968.

11. The Civilian Industrial Technology Program: A History and Evaluation. Jerome Yurow and F. Cecil Brenner. National Bureau of Standards Report 10,182. November 1969.

12. Results of an Experimental Program to Provide Low Cost Computer Searches of the NASA Information File to University Graduate Students in the Southeast. D. M. Phillips and F. O. Smetana. NASA Cr-95724. June 1968.

13. Appendicies to a Final Report to NSF. A. D. Little, Inc. 1963. U. S. Clearinghouse Document P.B. 181,548 A.

APPENDIX I

Commands Used During Study Trials

Command	Total
X = KTA	78
X = KTA + KTB	12
X = KTA + KTB + KTC	4
X = KTA + KTB + KTC + KTD	1
X = KTA + KTB + KTC + KTD + KTE	1
X = KTA * KTB	5
X = KTA * KTB * KTC	3
X = ·A + ·B	1
X = ·A * ·B	41
X = ·A * ·B * ·C	4
X = ·A : ·B	1
X = ·A + KTB	8
X = ·A + KTB + KTC	5
X = ·A * KTB	3
X = ·A · KTB	2
UPRINT X	44
PRPOST (X)	29
PRPOST (X) T,A,J,Y,P,CT (or any set of these)	17
PRINT X title name (with or without "no dup")	5
Classify X by doc, or author, or class	21
Thesaurus	7
Fldfrg X no dup.	7
Citations X	3
X = KTA/1/name	1

APPENDIX II

Statistics on Trials Involving

Problems Suggested by Representatives of Industry

	Average	Range
A. General		
Computer time used per problem in seconds.	~ 210	72.3-529.5
Elapsed phone time used per problem in minutes	~ 21	4.6 - 56.7
Ratio of elapsed to computer time	~ 6 to 1	4-10 to 1
B. Computer time in seconds to secure an intersection between files A and B on the command, X = ·A * ·B	1.48	0.7 - 4.6
C. Computer time in seconds used in response to the command, PRPOST (document number)	29.8	14.9 - 45.2

D. Breakdown of responses to the command, PRPOST (document number) TA,A,J,Y,P,CT (or any combination thereof)

Number of Items of Information Requested	Items Secured*	Time in Seconds
1	1	28.8
2	2	29.0
2	2	28.9
2	1	39.3
2	2	26.2
3	3	32.5
3	2	13.0
3	2	13.9
3	2	20.2
3	1	22.0
4	4	19.4
4	4	32.3
4	4	23.5

(Continued)

APPENDIX II
(cont.)

D. (continued)

Number of Items of Information Requested	Items Secured*	Time in Seconds
4	4	11.1
4	2	24.5
5	2	30.9
5	2	20.2
	Average ~	24.5

* The discrepancy between "requested" and "secured" resulted from situations where the document, for example, was found to be a patent, whereas it had been incorrectly assumed to be a publication in a journal.

APPENDIX III

FIGURES AND COMPUTER PRINTOUTS

A. Computer time in seconds required to make a search for a keyterm as related to the number of documents found carrying that keyterm, using the command **X** = keyterm A. See Figure 1.

B. Computer time in seconds required to respond to the command, UPRINT **X**, where **X** was a file, the contents of which had been accumulated and contained from 1 to 23 documents. See Figure 2.

C. Printouts of the TIRP and STRC responses to a search having to do with documents concerned with, "The analyzing and especially the chemical analysis of caprolactam, polycaprolactam and nylon-6 for impurities". See attachments.

APPENDIX III - A

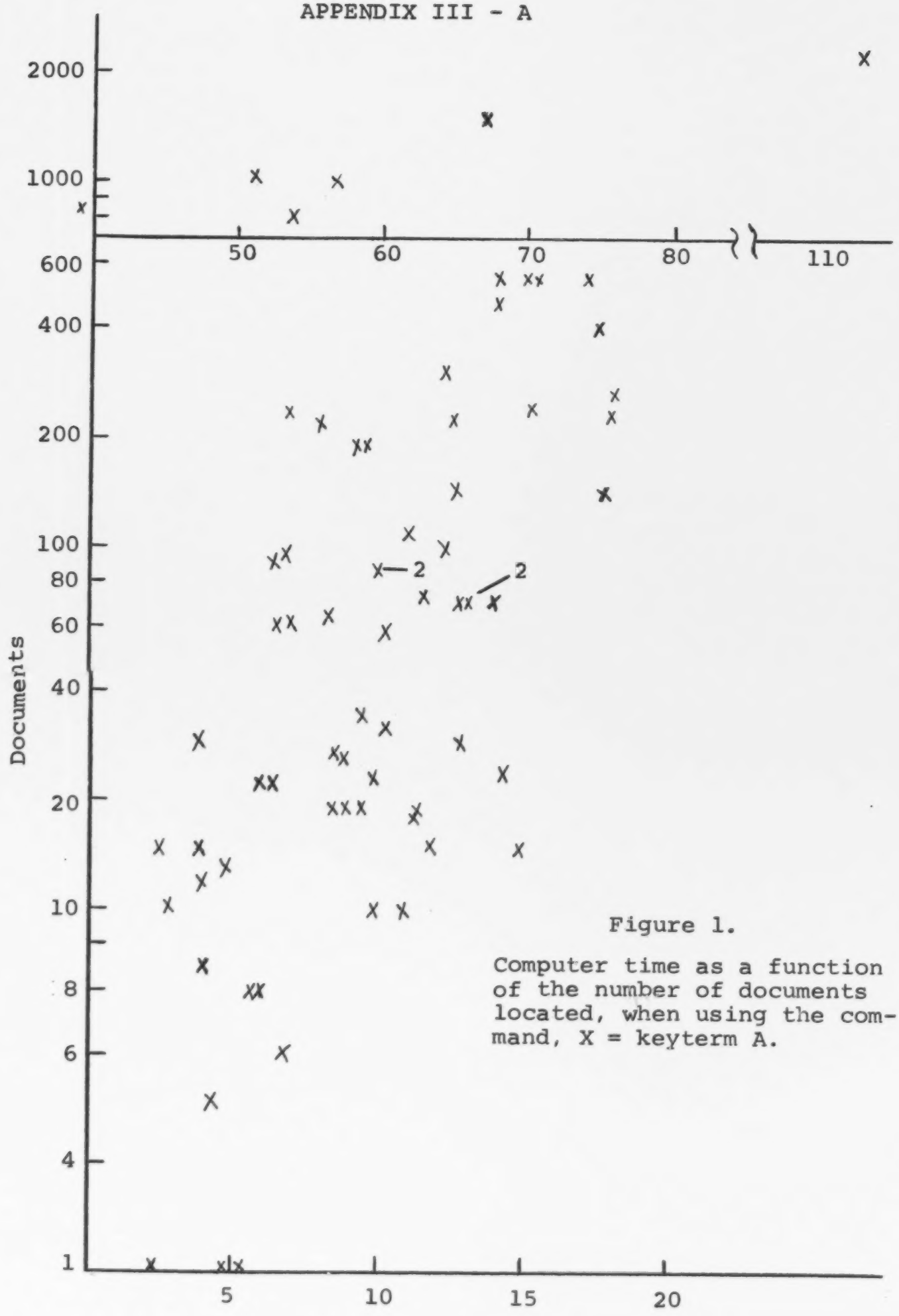


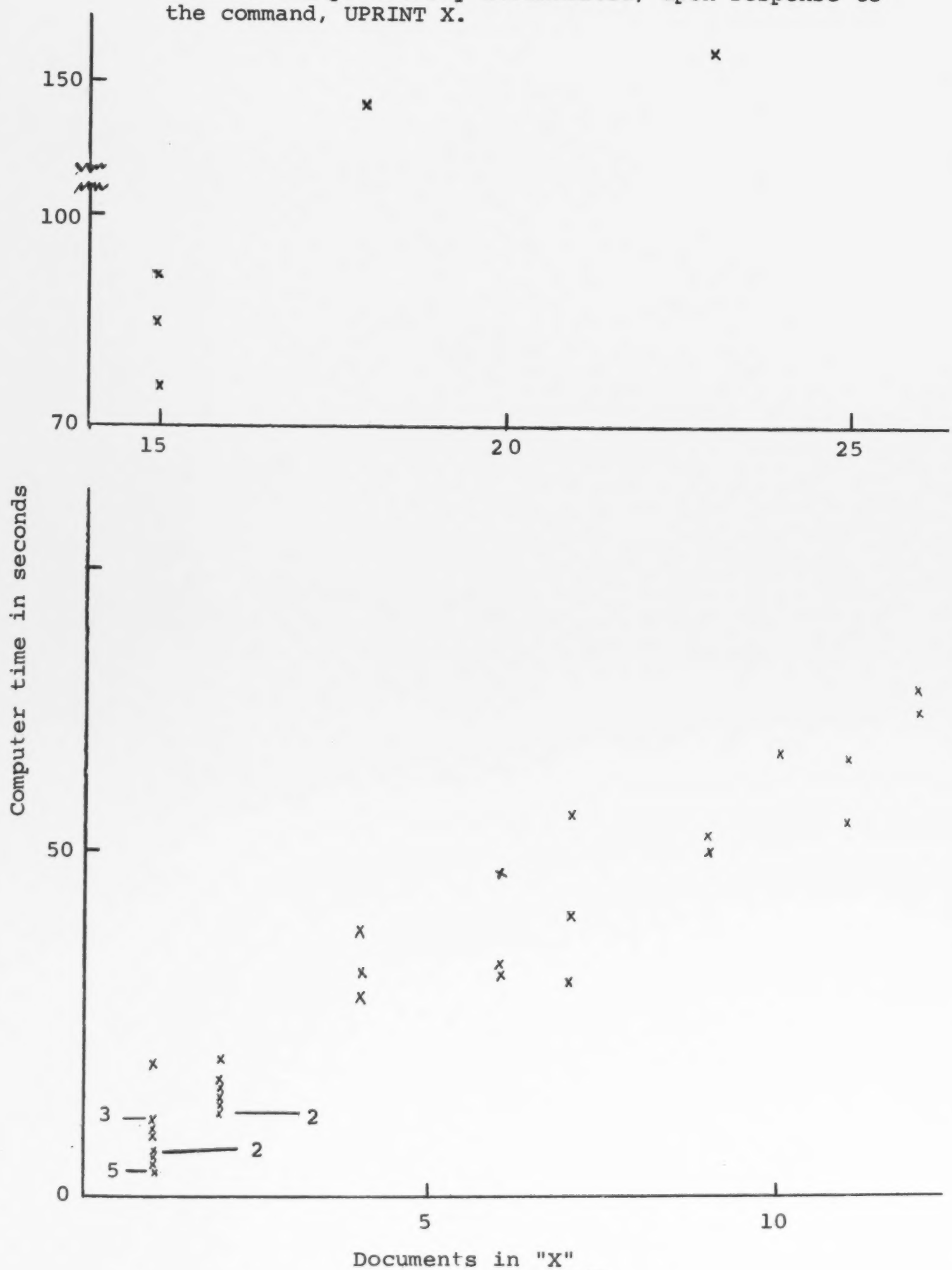
Figure 1.

Computer time as a function of the number of documents located, when using the command, X = keyterm A.

APPENDIX III - B

Figure 2.

Computer time as a function of the number of documents in a file X previously accumulated, upon response to the command, UPRINT X.



APPENDIX III - C

TIRP Printout
(Copied)

```
READY .
A = IMPURITIES

W 1616.5

A HAS  33 PENTADS,  26 DOCUMENTS.

R 8.6 + 8.6

B = .A+CHEMICAL ANALYSIS+ANALYZING

W 1617.2

B HAS  732 PENTADS,  393 DOCUMENTS.

R 45.2 + 36.6

C=CAPROLACTAM+POLYCAPROLACTAM+NYLON 6

W 1619.1

C HAS  249 PENTADS,  109 DOCUMENTS.

R 69.1 + 23.9

BC=.B*.C

W 1620.3

BC HAS  28 PENTADS,  7 DOCUMENTS.

R  80.9 + 11.7

UPRINT BC

1 . HAS   7 PENTADS,  7 DOCUMENTS.

64-D10905,02  U
      DYE ABSORPTION IN THE CONTINUOUS DYEING OF POLYAMIDE
      FABRICS.
      KORCHAGIN, M. V.
```

64-D61508,02
BENSOCYLATION OF POLYAMIDE FIBRES.
JANSEH, M.

64-D61509,02
MODIFICATION OF PROPERTIES OF POLYCAPROLACTAM FIBRE MATERIALS BYBENZOYLATION.
JANSEH, M.

65-D24007,04
YELLOWING AND DEGRADATION OF E-CAPROLACTAM AND POLYCAPROLACTAM. I - YELLOWING OF CAPROLACTAM BY OXYGEN.
ROTH, W.

66-D24401,02
WOOL SOLUBILITY REAGENTS.
GRUNDEA, M.

66-D36602,02 U
GRAFT POLYMERISATION OF ACRYLONITRILE ON NYLON.
GLUKHOV, V. I.

67-D43014,02 U
APPLICATION OF DIFFERENTIAL THERMAL ANALYSIS IN TEXTILE CHEMISTRY II. - DETERMINATION OF FINE STRUCTURE IN NYLON 6 AND POLYESTER FIBERS BY CALORIMETRIC METHODS.
JACOBASCH, H. J.

THERE ARE 7 ENTRIES IN 1.

R 135.0 + 54.1

STRC Printout

(Copied)

QUESTION 1 CONTAINS 6 TERMS AND 2 GROUPS

TOTAL NUMBER OF POSTINGS FOR THIS QUESTION = 5230

GROUP 1 CONTAINS 3 TERMS

IMPURITIES OR 6 CHEMICAL ANALYSIS OR 6
ANALYZING AND 6
*** AND ***

GROUP 2 CONTAINS 3 TERMS

CAPROLACTAM OR 6 POLYCAPROLACTAM OR 6
NYLON 6
*** ***

THE TERM IMPURITIES HAS BEEN READ IN

THE TERM CHEMICAL ANALYSIS WILL BE ADDED TO THE UNION
THERE ARE 368 HITS IN THIS UNION

THE TERM ANALYZING WILL BE ADDED TO THE UNION
THERE ARE 403 HITS IN THIS UNION

THE RESULTS OF GROUP 1 FOLLOW
THERE ARE 403 HITS IN GROUP 1

THE TERM CAPROLACTAM HAS BEEN READ IN

THE TERM POLYCAPROLACTAM WILL BE ADDED TO THE UNION
THERE ARE 25 HITS IN THIS UNION

THE TERM NYLON 6 WILL BE ADDED TO THE UNION
THERE ARE 109 HITS IN THIS UNION

THE RESULTS OF GROUP 2 FOLLOW

1344	1651	1784	1824
1873	1914	1919	1973
1976	2035	2096	2185
2193	2730	2763	2937
3166	3445	3476	3586
3815	3852	3856	3889
3954	3968	4358	4413
4567	4599	4748	4769
4820	4921	4937	4979
5025	5157	5297	5324
5392	5528	5532	5539
5540	5604	5666	5740
5759	5811	5823	5858
5960	5993	6072	6107
6198	6364	6443	6447
6797	6799	6854	6903
6954	7018	7096	7153
7219	7225	7480	7530
7532	7725	7831	7876
7968	7985	8027	8032
8069	8173	8344	8587

(Continued)

8657	8688	8743	8758
8856	8983	9129	9172
9303	9347	9388	9414
9474	9663	9685	9840
9966	9976	10093	10114
10122	10161	10215	10389
10399	0	0	0

THERE ARE 109 HITS IN GROUP 2

GROUP 2 WILL INTERSECT WITH GROUP 1 (OR THE PREVIOUS INTERSECTION)

THIS IS THE FINAL HIT LIST

5297	5858	5993	7219
7725	7968	9840	0

THERE ARE 7 HITS IN THIS INTERSECTION

THIS SEARCH WILL BE SAVED WITH IDENTIFYING KEY NO. 205570 205

STRC-IVS BIBLIOGRAPHIC FILE - MIT
NORTH CAROLINA SCIENCE AND TECHNOLOGY RESEARCH CENTER

64-D109-05,02
A VILYENSKAYA, B. M.
KORCHAGIN, M. V.
T DYE ABSORPTION IN THE CONTINUOUS DYEING OF POLYAMIDE FABRICS.
J TEKSTIL.PROM.
Y 1963
V NO. 10, PP. 8-13.
05297 STRC ACCESSION NUMBER

64-D615-08,02
A JANSEH, M.
T BENSOYLATION OF POLYAMIDE FIBRES.
J FASERFORSCH. UND TEXTILTECH.
Y 1964
V V. 15, PP. 372-380.
05858 STRC ACCESSION NUMBER

64-D615-09,02
A VON HORNUFF, G.
JANSEH, M.

T MODIFICATION OF PROPERTIES OF POLYCAPROLACTAM FIBRE MATERIALS BYBENZOYLATION.
J MELLIAND TEXTILBER.
Y 1964
V V. 45, PP. 768-776.
05993 STRC ACCESSION NUMBER

65-D240-07,04
A ROTH, W.
SCHROTH, R.
T YELLOWING AND DEGRADATION OF E-CAPROLACTAM AND POLYCAPROLACTAM. I - YELLOWING OF CAPROLACTAM BY OXYGEN.
J FASERFORSCH. UND TEXTILTECH.
Y 1965
V V. 16, PP. 37-41.
07219 STRC ACCESSION NUMBER.

66-D244-01, 02
A GRUNDEA, M.
IFRIM, S.
T WOOL SOLUBILITY REAGENTS.
L ENGLISH
J INDUSTRIA TEXTILA AND ABS. RUMAN. TECH. LIT.
Y 1964 AND 1965
V V. 15, PP. 671-672, AND V. 1, P. 886.
07725 STRC ACCESSION NUMBER

66-D366-02,02
A KURILENKO, A. I.
GLUKHOV, V. I.
T GRAFTPOLYMERISATION OF ACRYLONITRILE ON NYLON.
J DOKLADY AKAD. NAUK, S. S. S. R.
Y 1966
V V. 166, NO. 4, PP. 901-904
07968 STRC ACCESSION NUMBER

67-D430-14,02
A VON HORNUFF, G.
JACOBASCH, H. J.
T APPLICATION OF DIFFERENTIAL THERMAL ANALYSIS IN TEXTILE CHEMISTRY II. - DETERMINATION OF FINE STRUCTURE IN NYLON 6 AND POLYESTER FIBERS BY CALORIMETRIC METHODS.
J FASERFORSCH. UND TEXTILTECH.
Y 1967
V V. 18, PP. 282-288
09840 STRC ACCESSION NUMBER